Inferring trust (CLIMA 2005)

Mehdi Dastani, Andreas Herzig, Joris Hulstijn, Leendert W. N. van Der Torre

Inferring Trust

Mehdi Dastani[1], Andreas Herzig[2], Joris Hulstijn[3], and Leendert van der Torre[4]

[1] Utrecht University, The Netherlands, mehdi@cs.uu.nl
[2] IRIT, Toulouse, France, herzig@irit.fr
[3] Vrije Universiteit, Amsterdam, The Netherlands jhulstijn@feweb.vu.nl
[4] CWI, Amsterdam, The Netherlands torre@cwi.nl

Abstract. In this paper we discuss Liau's logic of Belief, Inform and Trust (BIT), which captures the use of trust to infer beliefs from acquired information. However, the logic does not capture the derivation of trust from other notions. We therefore suggest the following two extensions. First, like Liau we observe that trust in information from an agent depends on the topic of the information. We extend BIT with a formalization of topics which are used to infer trust in a proposition from trust in another proposition, if both propositions have the same topics. Second, for many applications, communication primitives other than inform are required. We extend BIT with questions, and discuss the relationship with belief, inform and trust. An answer to a question can lead to trust, when the answer conforms to the beliefs of the agent.

1 INTRODUCTION

Trust is an issue which emerges in many subareas of artificial intelligence, in particular in multiagent systems, reputation systems, e-institutions, and electronic commerce [1]. Liau [2] proposes an elegant, simple, but expressive modal logic as an extension of multi-agent epistemic logic. The three main ingredients are modal operators for belief (B), inform (I), and trust (T). The central axiom expresses that if an agent trusts another agent with respect to a proposition, and it has been informed by that agent that the proposition is true, then it believes that proposition.

The logic explains the consequences of trust, but it does not explain where trust comes from. The only optional axiom discussed by Liau that derives positive trust formulas is so-called transferability, which says that trust in one agent can lead to trust in another agent with respect to the same proposition. In this paper, we study two other ways in which trust can be derived. We do this by first enriching Liau's framework with topics and questions, and then by investigating the following issues.

1. How to use topics to infer trust? Like Liau we observe that trust in information depends on the topic of the information. We extend BIT with a formalization of topics. Topics can be used to infer trust in a proposition from trust in another proposition, if both propositions have the same topics.
2. How to use communication to infer trust? For many applications, communication primitives other than inform are required. We extend BIT with questions and discuss the relationship with belief, inform and trust. An answer to a question can also lead to trust, when an agent tests another agent by questioning him and the answer conforms to the beliefs of the agent.

We formalize topics and questions in terms of non-normal modal operators. To obtain a simple axiomatization of our semantically defined operators we re-formalize them in terms of operators from normal modal logic using a technique known as *simulation*. Moreover, Liau uses a non-normal modal logic to formalize trust, i.e., his notion of trust is not closed under tautologies, nor under conjunction nor implication: agent i does not necessarily trust that $\top$, trust that $\varphi \wedge \psi$ does not imply trust that φ, and validity of $\varphi \supset \psi$ does not entail that trust that φ implies trust that ψ. In order to work in a uniform and simple framework we also simulate the non-normal trust operator, using a combination of normal modal logic operators. The reductions or simulations use the fact that "normal modal logics can simulate all others" [3, 4].

The layout of this paper is as follows. In Section 2 we introduce the running example. In Section 3 we repeat and discuss Liau's BIT logic, and we formalize the running example in it. In Section 4 and 5 we introduce topics and questions, as well as the principles permitting to infer trust that can be based on them.

2 RUNNING EXAMPLE

We use the following example to motivate and illustrate our extensions of Liau's logic.

Agent i wants to know the interest rate, which is of vital importance for his portfolio management. He has found three web-services s_1, s_2 and s_3 that present financial information, but he does not know whether they deliver up to date information, or whether the information is correct at all. In other words, agent i does not know which web-service to trust. Suppose agent i knows the latest exchange rates for the euro against the dollar, and asks the web-services about this piece of information. If they do not provide the correct information, then the agent concludes that the web-services are not trustworthy. Otherwise, if they supply the correct exchange rate, then the agent trusts them with respect to financial information. Thus he then knows whom to ask about the interest rate, in order to use this piece of information in his portfolio management.[5]

In this paper, we ignore the dynamics and time aspects[6] involved in this example and discuss the formalization of three aspects of this example.

1. First we express the example in Liau's BIT logic. What can be said there is that
   - if the agent trusts the web-service, then he believes what he is being informed about;
   - if a web-service has informed the agent about something it believes to be false, then the agent does not trust the web-service.
2. To relate the question about exchange rates with the question about interest rates, we introduce the notion of topic. Both exchange and interest rates have the topic of financial information. So, when the web-service can be trusted on exchange rates,

[5] We assume that the web-service is not a strategic player, in the sense of Goffman's strategic interaction [5], that is, we assume that the web-service does not have something to gain by making you believe that it is trustworthy but not being so. In this sense this example is less complex than issues around trust found in electronic commerce.

[6] We do not discuss the state transitions based on communication actions such as inform and question.

it can be trusted on the whole topic of financial information, and therefore it can be trusted on interest rates.

3. Based on the hypothesis that in general agents are not being informed by a web-service by accident, but are being informed as the result of a question being submitted to the web-service, we extend the system with a question operator. An agent can then infer trust in a web-service, in case the web-service has informed the agent in accordance with the agent's current beliefs.

3 BIT

In this section we repeat and discuss Liau's logic BIT [2], and we formalize the running example in it. Definition 1 presents the language of the basic BIT logic, where $B_i\varphi$ is read as 'agent i believes φ', $I_{ij}\varphi$ as 'agent i acquires information φ from agent j', and $T_{ij}\varphi$ as 'agent i trusts the judgment of agent j on the truth of φ'. In the rest of this paper, we read $I_{ij}\varphi$ as 'agent i is being informed φ by agent j' or 'agent i has been informed φ by agent j'. For the purpose of this paper, these three readings can be regarded as synonymous.

Definition 1 (BIT language). *Assume we have n agents and a set Φ_0 of countably many atomic propositions. The well formed formulae of the logic BIT is the least set containing Φ_0 that is closed under the following formation rules:*

- *if φ is a wff, then so are $\neg\varphi$, $B_i\varphi$, $I_{ij}\varphi$ and $T_{ij}\varphi$ for all $1 \leq i \neq j \leq n$, and*
- *if φ and ψ are wffs, then so is $\varphi \vee \psi$.*

As usual, other classical boolean connectives are defined as abbreviations.

Definition 2 presents the axiomatic system for basic BIT. Beliefs are represented by a normal KD45 modal operator; inform by a normal KD modal operator, and trust by a non-normal modal operator.

Definition 2 (BIT). *The basic BIT logic contains the following axioms and is closed under the following set of inference rules:*

P *propositional tautologies*
B1 $[B_i\varphi \wedge B_i(\varphi \supset \psi)] \supset B_i\psi$
B2 $\neg B_i\bot$
B3 $B_i\varphi \supset B_iB_i\varphi$
B4 $\neg B_i\varphi \supset B_i\neg B_i\varphi$
I1 $[I_{ij}\varphi \wedge I_{ij}(\varphi \supset \psi)] \supset I_{ij}\psi$
I2 $\neg I_{ij}\bot$
C1 $(B_iI_{ij}\varphi \wedge T_{ij}\varphi) \supset B_i\varphi$
C2 $T_{ij}\varphi \supset B_iT_{ij}\varphi$
R1 (Modus Ponens, MP): from $\vdash \varphi$ *and* $\vdash \varphi \supset \psi$ *infer* $\vdash \psi$
R2 (Generalization, Gen): from $\vdash \varphi$ *infer* $\vdash B_i\varphi$ *and* $\vdash I_{ij}\varphi$
R3 from $\vdash \varphi \equiv \psi$ *infer* $\vdash T_{ij}\varphi \equiv T_{ij}\psi$

Liau discusses several possible extensions of the basic BIT logic: additional axiom C3 is called symmetric trust, C4 is called transferability, C5 is called cautious trust, and axiom C6 is called the ideal environment assumption.

C3 $T_{ij}\varphi \supset T_{ij}\neg\varphi$	(symmetric trust)
C4 $B_i T_{jk}\varphi \supset T_{ik}\varphi$	(transferability)
C5 $T_{ij}\varphi \supset B_i[(I_{ij}\varphi \supset B_j\varphi) \wedge (B_j\varphi \supset \varphi)]$	(cautious trust)
C6 $I_{ij}\varphi \equiv B_i I_{ij}\varphi$	(ideal environment)

To understand Liau's logic, first observe that an agent can trust another agent, without believing that the other agent is sincere and competent, as in other logics of trust, see for example [6]. This is expressed by the central axiom (C1), which is weaker than the inference from a combination of sincerity $I_{ij}\varphi \supset B_j\varphi$ and competence $B_j\varphi \supset \varphi$ by the trusted agent, which are the respective constituents of cautious trust in C5.

Secondly, observe that the logic is focussed on the formalization of consequences of trust, not on how trust is derived. That is, axiom C1 characterizes how trust in a proposition may lead to a belief in that proposition (in case of an inform), but little is said about the derivation of trust. Axiom C3 relates trust in a proposition to trust in its negation, and axiom C4 derives trust in an agent from trust in another agent. There are no axioms that derive trust from an inform, or that relate trust in a proposition to trust in another proposition, except for the negation in C3.

Thirdly, it should be observed that the fact that the trust operator is non-normal, means that using axiom C1 we can derive $B_{ij}\varphi$ from $B_{ij}I_{ij}(\varphi \wedge \psi)$ and $T_{ij}\varphi$, but we cannot derive $B_{ij}\varphi$ from $B_{ij}I_{ij}\varphi$ and $T_{ij}(\varphi \wedge \psi)$. There are good reasons for this, for which we refer to Liau's paper. Liau presents the following standard semantics for his logic. We do not mention the semantic constraints for the additional C3-C6.

Definition 3 (Semantics BIT). *A BIT model is a tuple*

$$\langle W, \pi, (B_i)_{1\leq i\leq n}, (I_{ij})_{1\leq i\neq j\leq n}, (T_{ij})_{1\leq i\neq j\leq n}\rangle$$

where W is a set of possible worlds, $\pi : \Phi_0 \to 2^W$ is a truth assignment mapping each atomic proposition to the set of worlds in which it is true, $(B_i)_{1\leq i\leq n} \subseteq W \times W$ are serial, transitive and Euclidian binary relations on W, $(I_{ij})_{1\leq i\neq j\leq n} \subseteq W \times W$ are serial binary relations on W, and $(T_{ij})_{1\leq i\neq j\leq n}$ are binary relations between W and the power set of W. Moreover, the satisfaction relation is defined as follows.

1. $M, w \models p$ *iff* $w \in \pi(p)$
2. $M, w \models \neg\varphi$ *iff* $M, w \not\models \varphi$
3. $M, w \models \varphi \vee \psi$ *iff* $M, w \models \varphi$ *or* $M, w \models \psi$
4. $M, w \models B_i\varphi$ *iff for all* $u \in B_i(w)$, $M, u \models \varphi$
5. $M, w \models I_{ij}\varphi$ *iff for all* $u \in I_{ij}(w)$, $M, u \models \varphi$
6. $M, w \models T_{ij}\varphi$ *iff* $|\varphi| = \{u \in W \mid M, u \models \varphi\} \in T_{ij}(w)$, *where $|\varphi|$ is called the truth set of φ.*

The corresponding constraints for axioms C1 and C2 are:

m1 *For all $S \in T_{ij}(w)$, if $(B_i \circ I_{ij})(w) \subseteq S$, then $B_i(w) \subseteq S$, where '$\circ$' denotes the composition operator between two binary operations;*

m2 $T_{ij}(w) = \cap_{u\in B_i(w)} T_{ij}(u)$.

The logic may seem relatively simple, but – although Liau does not discuss such applications – we can already use the logic to reason about relatively complex phenomena such as trust in the ignorance of agents $T_{ij}(\neg B_j\varphi \wedge \neg B_j\neg\varphi)$ or some aspects of trusted third parties $(B_iI_{ij}T_{jk}\varphi \wedge T_{ij}T_{jk}\varphi) \supset T_{ik}\varphi$.

The following example formalizes some aspects of the running example.

Example 1. Assume a finite set of atomic propositions $i(0.0), \ldots, i(10.0)$ denoting interest rates, and a finite set of atomic propositions $e(0.50), \ldots, e(2.00)$ denoting exchange rates, where the interval and step size are chosen arbitrarily. Moreover, let the set of agents be $\{i, s_1, s_2, s_3\}$. From axiom C1, by contraposition we have the following set of instances, for $s \in \{s_1, s_2, s_3\}$ and $r \in \{0.50, \ldots, 2.00\}$, which states that if an agent i believes that a web-service s has informed him about an exchange rate which i does not believe, then agent i will not trust that web-service.

$$B_iI_{is}e(r) \wedge \neg B_ie(r) \supset \neg T_{is}e(r)$$

Moreover, axiom C1 also implies the following set of instances, for $s \in \{s_1, s_2, s_3\}$ and $r \in \{0.0, \ldots, 10.0\}$, which states that if an agent i believes that the web-service s has informed him about the interest rates, and i trusts s, then agent i believes the interest rates.

$$B_iI_{is}i(r) \wedge T_{is}i(r) \supset B_ii(r)$$

Finally, if agent i trusts the web-service s with respect to some interest or exchange rates, then i also trusts s with respect to other rates. This can be 'hard-coded' with the following set of assumptions, for $s \in \{s_1, s_2, s_3\}$, $r_1, r_3 \in \{i(0.0), \ldots, i(10.0)\}$ and $r_2, r_4 \in \{e(0.50), \ldots, e(2.00)\}$.

$$T_{is}i(r_1) \vee T_{is}e(r_2) \supset T_{is}i(r_3) \wedge T_{is}e(r_4)$$

Hence Liau's logic already allows to infer new beliefs via trust, and to infer distrust. What it does not allow is to infer trust, which is what the rest of the paper is about.

4 Topics

For trust it matters what a formula "is about": its topic. Agents have a certain area of expertise or competence. If they are trustworthy on some formulas, then they are likely to be trustworthy on other formulas that have the same topic. That will lead to a principle of inference that, for example, trust in one financial rate implies trust in another financial rate. We formalize a principle of *topical trust*. Liau already recognizes the need for topical trust, as his third item for further research:

> "A special case of symmetric trust, called topical trust, is considered without standard axiomatization. This problem may be remedied by introducing the topics of propositions into the language. For example, in a logic of aboutness [7], a sorted binary predicate $A(t,\text{`}p\text{'})$ is used to denote "sentence '$p$' is about topic $t$". If our BIT language is extended with such a predicate, then we can formulate axioms as: $A(t,\text{`}\varphi\text{'}) \supset T_{ij}\varphi$ when j is specialized at topic t, or

more strongly, as $(A(t_1,\text{`}\varphi\text{'}) \vee \ldots \vee A(t_k,\text{`}\varphi\text{'})) \equiv T_{ij}\varphi$ when the set of topics at which an agent is specialized are $[t_1, \ldots, t_k]$. However, further research is needed to see how the semantics can be changed to accommodate this syntactic extension."

Our extension of BIT logic with topics is loosely inspired by a proposal of Herzig and Longin. Whereas Herzig and Longin formalize the notion of topics in the metalanguage, we will formalize it using standard normal modal operators.

4.1 Herzig and Longin

The conceptual model of Herzig and Longin [8] is visualized in Figure 1. It contains a meta theory with the following three relations:

- A competence function that relates agents to topics, namely those topics in which the agent is an expert.
- A subject function that relates propositions to topics, namely those topics that the propositions are about.
- A scope function that relates actions (such as inform) to topics. Actions which are affected by the topic of proposition are listed here.

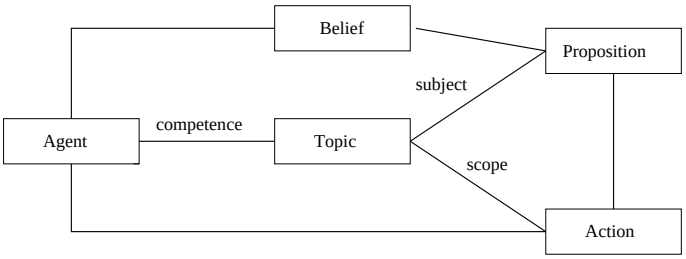

Fig. 1. Conceptual Model of Trust

These concepts enable one to formulate principles of belief update. Informally, they can be expressed as follows:

- If a formula φ holds, and an agent is informed about a proposition which does not share any topic with φ, then φ persists;
- If an agent j is competent on a topic and φ belongs to that topic, then an inform by agent j that φ implies belief that φ.

The first principle is not relevant for this paper, because the BIT logic only considers the state of the world at one moment. An extension with time is very interesting, but beyond the scope of this paper. The second principle implies that if an agent is competent on a proposition φ and all topics of proposition ψ are also topics of φ, then the agent is competent on ψ, too. It is the latter issue which we formalize in the BIT logic, simply

replacing belief in competence by trust. This move disregards the distinction between the two, in the sense that belief in someone's competence may lead to trust, but this need not always be the case and more importantly, trust can be based on other reasons than belief in competence. Note that both Demolombe and Herzig and Longin take a syntactic approach. Aboutness $A(t,\text{`}p\text{'})$ and 'subject' are relations between formulas and some set of objects $t_1, ..., t_n$ called topics with no additional structure. By contrast we handle topics in the semantics.

4.2 Simulation

In this section we formalize the trust and topic operators, using a technique called *simulation*. This means that – typically complex – operators are defined in terms of standard normal modal operators. For example, the simulation of the non-normal trust operator in normal modal logic means that the trust operator is defined using normal operators, but that the operator itself behaves like a non-normal operator.

The advantages of simulation are twofold. First, the advantage of classical simulations such as the simulation of various kinds of non-normal modal logics in [3, 4] is that theorem provers of normal modal logic can be used for proving theorems of non-normal modal logic. This advantage also holds for the simulation of the non-normal trust operator in normal modal logic. This means, among other things, that it becomes easier to have a theorem prover test specifications written in the extended BIT logic. Second, the advantage that motivates the simulation in this paper is that such a simulation gives us a direct axiomatization of the logic, which would not be obtained if the operators were only defined semantically. In that case, additional axioms would have to be given to characterize the semantic notions.

Consider the trust operator, which is a non-normal modal operator. This operator can be simulated using three standard normal modal operators $\Box^1_{ij}$, $\Box^2$ and $\Box^3$ [4]

$$T_{ij}\varphi \equiv \Diamond^1_{ij}(\Box^2\varphi \wedge \Box^3\neg\varphi)$$

where $\Diamond\varphi$ abbreviates $\neg\Box\neg\varphi$ as usual.

To understand the reduction remember that truth of $T_{ij}\varphi$ in a world w of a model M means that there is a truth set (neighborhood) $S \in T_{ij}(w)$ such that $M, w' \models \varphi$ for every $w' \in S$, and $M, w'' \not\models \varphi$ for every $w'' \notin S$. Thus $\Diamond^1_{ij}$ enables us to refer to the existence of a truth set (neighborhood), $\Box^2$ is used to express the truth of φ in S, and $\Box^3$ expresses the falsehood of φ outside S.

4.3 Topic as enumeration of options

In this paper, we assume that propositions have topics and that topics are shared by all agents[7]. For example, the proposition $i(5.0)$ has financial information as its topic. Moreover, in the Herzig-Longin approach propositions can belong to two or more topics, though this does not play a role in the example. Consequently, a complication of the formalization of topics is that we not only have to state which topics there are, but

[7] We assume here that topics are shared by all agents to simplify our presentation.

that these are all the topics available. It is only by making explicit all given topics, that we can quantify over topics. For this reason, we introduce both an operator topic and an operator all_topics. We identify a topic with the set of atomic propositions that have this topic as a subject (see above). For example, the topic financial information is identified with the set

$$\{i(0.0), \ldots, i(10.0), e(0.50), \ldots, e(2.00)\}$$

Such a topic set will be represented by a formula like

$$\mathsf{topic}(i(0.0) \times \ldots \times i(10.0) \times e(0.50) \times \ldots \times e(2.00))$$

in which '$\times$' is used to separate alternative options. Our encoding is as follows.

Definition 4 (Topics). *The language of BIT with topics is the language of BIT, together with clause*

- *if φ is a sentence of BIT, then so are $\Box^1\varphi$, $\Box^2\varphi$, $\Box^3\varphi$ and $\Box^4\varphi$.*

Moreover, we add the following abbreviations:

- $\varphi_1 \times \ldots \times \varphi_n \equiv \Diamond^2(\Box^3\varphi_1 \wedge \Box^4\neg\varphi_1) \wedge \ldots \wedge \Diamond^2(\Box^3\varphi_n \wedge \Box^4\neg\varphi_n) \wedge \Box^2((\Box^3\varphi_1 \wedge \Box^4\neg\varphi_1) \vee \ldots \vee (\Box^3\varphi_n \wedge \Box^4\neg\varphi_n))$
- $\mathsf{topic}(\varphi_1 \times \ldots \times \varphi_n) \equiv \Diamond^1(\varphi_1 \times \ldots \times \varphi_n)$
- $\mathsf{all\_topics}((\varphi_{1,1} \times \ldots \times \varphi_{1,n}); \ldots; (\varphi_{k,1} \times \ldots \times \varphi_{k,m})) \equiv \Box^1((\varphi_{1,1} \times \ldots \times \varphi_{1,n}) \vee \ldots \vee (\varphi_{k,1} \times \ldots \times \varphi_{k,m}))$
- $\mathsf{topic\_contained}(\varphi, \psi) \equiv \Box^1(\Diamond^2(\Box^3\varphi \wedge \Box^4\neg\varphi) \supset \Diamond^2(\Box^3\psi \wedge \Box^4\neg\psi))$

The topic notation with $\times$ may be read as a representation of a set. That is, due to the properties of the modal logic we have for example that $p \times q \times r$ implies $q \times p \times r$ or $p \times p \times q \times r$, but it does not imply for example $p \times q$.

The operator topic represents the set of propositions having the same topic; all_topics states furthermore that these are all topics available, and topic_contained formalizes the fact that all topics of the first element are also a topic of the second element. In our example $\mathsf{topic\_contained}(i(1.0), e(2.00))$ holds. In example 2 an explanation is given. So $\mathsf{topic\_contained}(\varphi, \psi)$ expresses that for every ($\Box^1$) topic, if formula φ has that topic ($\Diamond^2(\Box^3\varphi \wedge \Box^4\neg\varphi)$), then formula ψ has that topic too. It is the latter abbreviation which will be used to formulate a topic-based trust inference principle.

We assume that topics are treated as axioms, in the sense that they are known by all agents, and distribute over inform and trust operators. We therefore accept the following principles:

$$\mathsf{topic}(\varphi_1 \times \ldots \times \varphi_n) \equiv B_i\mathsf{topic}(\varphi_1 \times \ldots \times \varphi_n)$$
$$\mathsf{topic}(\varphi_1 \times \ldots \times \varphi_n) \equiv I_{ij}\mathsf{topic}(\varphi_1 \times \ldots \times \varphi_n)$$
$$\mathsf{topic}(\varphi_1 \times \ldots \times \varphi_n) \equiv T_{ij}\mathsf{topic}(\varphi_1 \times \ldots \times \varphi_n)$$

The semantics of BIT with topics extends the semantics of BIT with four binary accessibility relations that correspond to $\Box^1$ to $\Box^4$, that are interpreted in the usual way. The distribution of topic operators over the BIT modalities is characterized by the fact that in each world, the relevant accessibility relations are the same. Due to space limitations we do not give the details.

It may seem that our encoding of the topic operators is rather complicated, compared to for example [7], but the advantage is that we have a standard semantics. Moreover, an important advantage is that we can use the same methodology for questions too (see section 5).

4.4 Comparison with Janin and Walukiewicz

The encoding of the topic operator is a further extension of the simulation of non-normal modal operators mentioned above. This extension can be understood by analogy to work by Janin and Walukiewicz [9]. They define $a \rightarrow S =_{def} \bigwedge_{\varphi \in S} \Diamond^a \varphi \wedge \Box^a \bigvee_{\varphi \in S} \varphi$, where a is an index of a modal operator and S is set of formulas [9]. It means that world w satisfies formula $a \rightarrow S$ when any formula of S is satisfied by at least one a-successor of w, and all a-successors of w satisfy at least one formula of S. Classical modal operators are written as $\Diamond^a \varphi \equiv a \rightarrow \{\varphi, \top\}$ and $\Box^a p \equiv a \rightarrow \{\varphi\} \vee a \rightarrow \emptyset$. This is essentially the definition of bisimulation,[8] so the representation reflects the essence of modal logic. As we indicated above, we use the $\times$-notation instead of sets, so $S = \{p, q, r\}$ is represented by $p \times q \times r$. Like sets we have iteration and associativity, i.e., we can derive for example $p \times q \times q \times r$. However, also note that if modalities $\Box^a$ and $\Diamond^a$ are normal, then we can derive weakening: $(p \wedge q) \times r \rightarrow p \times r$. Since we do not like this property for topics, we use non-normal modal operators – to be precise, non-monotonic ones – that do not satisfy weakening. So, in our reduction of topics, we combine two ideas:

(a) $\Box \varphi \equiv \Diamond^2(\Box^3 \varphi \wedge \Box^4 \neg \varphi)$ (simulation, as before)
(b) $a \rightarrow S \equiv \bigwedge_{\varphi \in S} \Diamond^a \varphi \wedge \Box^a \bigvee_{\varphi \in S} \varphi$ (Janin and Walukiewicz)

These are combined using the definition of modality 2 according to (b), substituting $(\Box^3 \varphi \wedge \Box^4 \neg \varphi)$ for φ and substituting '2' for a, which gives us $\bigwedge_{\varphi \in S} \Diamond^2(\Box^3 \varphi \wedge \Box^4 \neg \varphi) \wedge \Box^2 \bigvee_{\varphi \in S} (\Box^3 \varphi \wedge \Box^4 \neg \varphi)$, which corresponds to the topic definition above. Since this only defines one topic, we still have to represent that "there is a topic", for which we use $\Diamond^1$.

4.5 Topics and trust

Now we can formalize the intuition that if a proposition is trusted, then also all other propositions are trusted which are based on the same topics. We call it *topic-based trust transfer* ($T3$).

$$\Diamond^1(\Diamond^2(\Box^3 \varphi \wedge \Box^4 \neg \varphi)) \wedge \mathsf{topic\_contained}(\varphi, \psi) \supset (T_{ij} \varphi \supset T_{ij} \psi) \qquad (T3)$$

We formalize the running example with topics. Since there is only one topic, the example is relatively simple.

Example 2. The topic financial information (f) is defined as follows.

$$f \equiv (i(0.0) \times \ldots \times i(10.0) \times e(0.50) \times \ldots \times e(2.00)) \qquad \mathsf{topic}(f) \qquad \mathsf{all\_topics}(f)$$

[8] This insight is attributed to Alexandru Baltag by Yde Venema.

In the first treatment of the example, the trust inference was 'hard coded'. Now, we use axiom $T3$ to derive: $T_{is}i(r_1) \vee T_{is}e(r_2) \supset (T_{is}i(r_3) \wedge T_{is}e(r_4))$. In particular, from $\mathsf{topic}(f)$ we can derive $\Diamond^1(\Diamond^2(\Box^3 i(r_1) \wedge \Box^4 \neg i(r_1)))$ and from $\mathsf{topic}(f)$ and $\mathsf{all\_topics}(f)$ we can infer $\mathsf{topic\_contained}(i(r_1), i(r_3))$. Using axiom $T3$, we can infer $T_{is}i(r_1) \supset T_{is}i(r_3)$. Similarly, we can infer $T_{is}i(r_1) \supset T_{is}e(r_4)$ and therefore $T_{is}i(r_1) \vee T_{is}e(r_2) \supset T_{is}i(r_3) \wedge T_{is}e(r_4)$. So the property that was postulated in Example 1, is now derived from our topic construction.

Finally, we note that Liau does not discuss the possibility to add $(T_{ij}\varphi \wedge T_{ij}\psi) \supset T_{ij}(\varphi \vee \psi)$, which at first hand looks reasonable, in particular when φ and ψ belong to the same topics. Such an axiom can be formalized with our topics. Also, by contraposition we can derive $\mathsf{topic\_contained}(\varphi, \psi) \supset (\neg T_{ij}\psi \supset \neg T_{ij}\varphi)$. In other words, if all topics of φ are a topic of ψ, distrust in ψ transfers to distrust in φ.

5 QUESTIONS

In this section, the logic of Liau is extended with questions, because of their specific relation to trust. Questions have been studied extensively as part of the semantics of natural language. In this paper we use the semantics of questions and answers of Groenendijk and Stokhof [10]. The idea is as follows. Conceptually, a question expresses a 'gap' in the information of the asker, to be filled by an answer of the right type. For example, a 'when'-question asks for a time or date. So a question specifies what its possible answers are. In the semantics, that means that a question separates the set of possible worlds into disjoint subsets, each of which correspond to a complete answer to the question. The resulting structure is a partition [10]. Technically, a partition is equivalent to an equivalence relation, called an *indistinguishability relation*: the agent does not distinguish between worlds that satisfy the same answer to a question. For a yes/no question there are two sets of worlds in the partition: worlds that correspond to the answer "yes", and worlds that correspond to the answer "no". For an alternative question like "Which color is the traffic light?", the partition corresponds to three possible answers: "red", "yellow" and "green". For an open question like "Who are coming to the party?", which asks about groups of people coming to the party, we would get possible answers ranging from "Nobody will come", "John will come", "Mary will come" and "John and Mary will come", up to ""Everybody will come". In other words, open questions are treated as alternative questions, where each selection from a contextually relevant set corresponds to one alternative.

Like in the case of topics, this conceptualization of questions can be encoded using the symbol '$\times$' to separate alternatives. We denote a question by an expression $\mathsf{question}_{ij}(\varphi_1 \times \ldots \times \varphi_n)$, where $\varphi_1 ... \varphi_n$ are the alternative answers. For example, "Which color is the traffic light?" is encoded by $\mathsf{question}_{ij}(\text{`}\textit{traffic\_light\_is\_red} \times \textit{traffic\_light\_is\_yellow} \times \textit{traffic\_light\_is\_green})$. Note that yes/no questions are a special case of alternative questions.

In some of the trust derivation cases, we need to express the fact that a possible answer was, either explicitly or implicitly, asked for. We use the Q_{ij}-operator for this. Expression $Q_{ij}\varphi$ means that agent i has posed a question to agent j for which φ is a

possible answer. In other words, $Q_{ij}\varphi$ holds in case $\mathsf{question}_{ij}(\psi_1 \times ... \times \psi_n)$ has been explicitly or implicitly posed by agent i to agent j, for $\varphi \equiv \psi_k$ and $1 \leq k \leq n$.

Definition 5 (Questions). *The language of BIT with topics and questions, is the language of BIT with topics, together with the following clause:*

- *if φ is a sentence of BIT with topics, then so is $\Box_{ij}\varphi$, for $1 \leq i \neq j \leq n$.*

Moreover, we add the following abbreviations:

- $\mathsf{question}_{ij}(\varphi_1 \times \ldots \times \varphi_n) = \Diamond_{ij}(\varphi_1 \times \ldots \times \varphi_n)$
- $Q_{ij}\varphi = \Diamond_{ij}\Diamond^2(\Box^3\varphi \wedge \Box^4\neg\varphi)$

The definition is analogous to the simulation of topics by a range of normal modal operators. The semantics of the BIT logic with topics and questions, extends the semantics of the BIT logic with topics, with a suitable accessibility relation corresponding to $\Box_{ij}$. In the semantics $\Diamond_{ij}$ or equivalently $\mathsf{question}_{ij}$ expresses the existence of a neighborhood corresponding to the answers to a question from agent i to j. The operators $\Diamond^2$ and $\Box^3$, $\Box^4$ are again used to express the properties of the $\times$-notation for alternatives. Note that like trust, but unlike topics, the semantics of questions is made relative to agents i and j. This expresses the intuition that topics are part of the general logical language, which is shared by all agents, whereas the questions that have been asked are particular for specific agents.

In a way, this provides only a minimal semantics. It does not express Groenendijk and Stokhof's idea of a partition. In case we want to model that answers to a question must be exclusive, and that the presented answers cover the whole logical space, i.e., that a question partitions the logical space, then we add the following axioms:

$$\mathsf{question}_{ij}(\varphi_1 \times \ldots \times \varphi_n) \supset (\varphi_i \wedge \varphi_j \supset \bot), \text{ for all } 1 \leq i \neq j \leq n$$
$$\mathsf{question}_{ij}(\varphi_1 \times \ldots \times \varphi_n) \supset (\varphi_1 \vee \ldots \vee \varphi_n \equiv \top)$$

5.1 Questions and Trust

The specific relation between questions and trust that we like to formalize in this section is based on the following intuition. If agent i has deliberately posed a question to an agent j to which agent i already believes the answer, and agent j has provided information that corresponds to the initial beliefs of agent i, then agent i will trust the second agent j. Otherwise, if agent j has provided the wrong answer, i.e. the information does not correspond to i's initial beliefs, then agent i will not trust agent j. This intuition is formalized by the following axioms which we call *question-based trust derivation* and *question-based distrust derivation* respectively.

$$(Q_{ij}\varphi \wedge B_i\varphi \wedge B_iI_{ij}\varphi) \supset T_{ij}\varphi$$
$$(Q_{ij}\varphi \wedge B_i\neg\varphi \wedge B_iI_{ij}\varphi) \supset \neg T_{ij}\varphi$$

Here, the combination of $Q_{ij}\varphi$ and $B_iI_{ij}\varphi$ is meant to express that $I_{ij}\varphi$ is a relevant response of agent j to a question posed by agent i. This reading may be problematic for a setting in which different questions can be posed, with the same kinds of answers.

For example an answer “at five” may be relevant to both “When does the bus come?” and ‘When does the train come?”. However, these problems are not essential for the phenomenon of inferring trust.

Using these axioms, we can formalize our running example.

Example 3. Agent i asks a web-service s the exchange rate: $\mathsf{question}_{is}(e(0.50) \times \ldots \times e(2.00))$ which implies $Q_{is}e(0.50) \wedge \ldots \wedge Q_{is}e(2.00)$. If the agent believes for example that the exchange is 1, $B_i e(1)$, and the web-service gives the correct answer, i.e., $B_i I_{is} e(1)$, then using the question-based trust creation axiom we can derive $T_{is}e(1)$. Similarly, in case the agent’s beliefs do not correspond to the answer, for example $B_i e(5)$ and therefore $B_i \neg e(1)$ because exchange rates are unique, we derive $\neg T_{is}e(1)$ by question-based distrust creation.

5.2 Questions and Topics

Questions turn out to be very similar to topics. In the example, the topic ‘financial information’ corresponds to a combination of the questions “What is the current interest rate?” and “What is the current exchange rate?”. In natural language semantics, relations between topics and questions have long been known. Van Kuppevelt [11] even defines topics in terms of the questions that are currently under discussion. By asking a question, the asker can manipulate the current topic of the conversation. As we noted above, topics are the same for all worlds and all agents. By contrast, we can use Q_{ij} to express the particular ‘questions under discussion’ for agents i and j. Under such an interpretation, it would make sense that questions were closed under topic: $Q_{ij}\varphi \wedge \mathsf{topic\_contained}(\varphi, \psi) \supset Q_{ij}\psi$. However, under such an implicit ‘questions under discussion’ interpretation, the question operator cannot be used to model that an agent explicitly asked for some information. But this is exactly the interpretation we need in the running example. We therefore use an intermediate step, first using question-based trust creation, and then applying the topic-based trust transfer principle.

Example 4. We would like to prove the following.

$$(B_i e(r) \wedge \mathsf{question}_{is}(\ldots \times e(r) \times \ldots) \wedge I_{is}e(r) \wedge \mathsf{topic\_contained}(e(r), i(r')) \wedge I_{is} i(r')) \supset B_i i(r')$$

Suppose $(B_i e(r) \wedge \mathsf{question}_{is}(\ldots \times e(r) \times \ldots) \wedge I_{is}e(r) \wedge \mathsf{topic\_contained}(e(r), i(r')) \wedge I_{is} i(r'))$. First, derive $Q_{is}e(r)$ by the definition of Q_{ij}, and subsequently $T_{is}e(r)$ by the principle of question-based trust creation. Second, derive $T_{is}i(r')$ from $(T_{is}e(r) \wedge \mathsf{topic\_contained}(e(r), i(r')))$ by topic-based trust transfer, and third, derive $B_i i(r')$ from $(I_{is}i(r') \wedge T_{is}i(r'))$ by Liau’s trust-based belief creation axiom C1. From these three formulas the desired implication can be obtained by principles of classical logic.

6 Further Research

6.1 Other communicative primitives

Suppose communication primitives $\mathsf{proposal}_{ij}\varphi$ and $\mathsf{request}_{ij}\varphi$ we added to the logic, to express that agent i received a proposal or request from j. Like an inform, an agent

will only accept a proposal when it trusts the agent's capabilities. And like a question, a request either indicates trust in the other agent's capabilities, or, analogous to our running example, a request is used to test the agent's capabilities. Once accepted, a proposal or request expresses a commitment of one of the participants to achieve some future state of affairs. Therefore we would have to further extend the logic with a 'see-to-it that' operator $E_i\varphi$ [12]. In that case, i's acceptance of a proposal by j can be expressed by an inform that i trusts the sender j to achieve the content of the proposal: $I_{ji}T_{ij}E_j\varphi$. Similarly, an acceptance of a request, is an inform that the accepter will achieve the content of the request: $I_{ji}E_i\varphi$. Thus in case of a proposal the sender will act upon acceptance, while in case of a request the receiver will act after having accepted.

$$\mathsf{proposal}_{ij}\varphi \wedge I_{ji}T_{ij}E_j\varphi \supset E_j\varphi$$
$$\mathsf{request}_{ij}\varphi \wedge I_{ji}E_i\varphi \supset E_i\varphi$$

6.2 Control Procedures

Trust can be based on personal relationships between agents, on a past experiences, or on a reputation that has been passed on by other trusted agents. In the absence of such direct trust in the other party, an agent has to rely on institutional control procedures to make sure that other agents will keep their part of the deal. Examples are banks to guarantee payment, or a bill of lading to guarantee shipping. However, if an agent does not understand a control mechanism, or does not trust the institutions that guarantee it, the mechanism is useless. Therefore one should also model trust in the control procedures. The general idea can be summarized as follows [1].

Transaction Trust = Party Trust + Control Trust

If we further analyze control trust, it comes down to two aspects. First, the agent must understand the workings of the control mechanism. For example, agent i understands that, within a shipment institution s, a bill of lading 'counts as' evidence of the goods having been shipped. A bill of lading is a specific kind of inform act. In BIT we write $I_{is}bill \supset I_{is}shipped$. Second, the agent must trust the institution s that guarantees the control mechanism. This can be expressed in BIT too: $T_{is}shipped$. Together, these rules implicate, that whenever the agent receives a bill of lading, it will trust that the goods have been shipped: $I_{is}bill \supset B_ishipped$. This translation is promising, but rather simplified. Further relations between Liau's BIT logic and evidential norms need to be investigated.

7 Related Research

The notion of trust has been studied extensively in the social sciences. For an overview of research on trust in the context of electronic commerce and multi-agent systems, see Tan and Thoen [1, 13]. Generally, trust is studied in relation to a transaction. Mayer et al. give the following definition of trust: "The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other party will perform a particular action important to the trustor, irrespective of the ability to monitor

or control that other party [14]". Note that risk is involved for the truster. A similar sentiment is found in the definition by Gambetta "Trust is the subjective probability by which an individual A expects that another individual B performs a given action on which its welfare depends" [15]. Both these definitions indicate that trust is subjective, and directed towards another agent. Trust reflects an interpersonal relation, that can be generalized to machines. This aspect is nicely reflected in the logic of Liau.

Aboutness and topicality have received a lot of attention in linguistics. A topic and its subtopics can be used to identify the structure of a text. For example, Grosz and Sidner [16] relate the topic of a discourse (also called *center* or *focus of attention*) to the intention that is intended to be conveyed by the author. More technical research on aboutness is done in the context of information retrieval [17]. Clearly, in information retrieval it matters under what circumstances we can say that two documents are "about the same topic".

A notion that is very similar to trust is found in the so called BAN logics [18], used to define authentication policies in computer security. Although there is no explicit notion of trust in these logics, sharing a secret key counts as a proof of being trusted. The primitives of BAN logic are as follows: i *sees* X, which means that agent i received a message containing X. This is similar to Liau's inform; j *said* X, which means that agent j actually sent a message containing X, and that in case j is to be trusted, X ought to be believed by i; i *controls* X, which can be interpreted as saying that agent i is trusted as an authority on X. This notion might be developed towards our use of topics. In BAN logics it is often used to represent trusted third parties, like authentication services; *fresh* X, which means that X has not been sent previously, and $i \overset{K}{\longleftrightarrow} j$, which means that agent i and j are entitled to use the same secret key K. Sharing a key counts as a proof of being trusted. There are several differences between BAN logics and Liau's BIT logic and the way they are used. An obvious difference is the use of keys, which is absent from Liau. Another difference concerns the perspective: Liau's logic takes the viewpoint of an individual agent: under what circumstances can I believe the content of a message? BAN takes the bird's eye view of a designer: how should I design my protocol to avoid secrets getting lost? The underlying logic is also different.

Finally, trust has been studied extensively in the context of a 'Grid'-like architecture for the sharing of resources and services [19]. Much of this work is applied. However, the underlying formal models that are developed in the context of such research [20] deserve to be compared with the BIT logic proposed here. Other formalizations in terms of modal logic also exist [21].

8 CONCLUSION

Trust plays an important role in advanced computer systems such as trust management systems in computer security [22] and reputation systems as used for example in eBay [23]. These applications define a much more precise notion of trust than the notion of trust used in social theories. Moreover, intelligent agents use trust mechanisms to reason about other agents, for example in cooperation, coordination, or electronic commerce. Agents that reason about their relations with other agents, such as agents reasoning about possible cooperation strategies, can benefit from reasoning about trust

explicitly. Liau's logic does not tell us much about the inner structure of trust, which may even be considered as a black box, but it does explain the relation between trust and other concepts, in particular the relation between trust, belief and information actions.

This paper presents two important extensions to Liau's BIT logic, which allow the derivation of trust. First, we extend the logic with topics. In this way, we can express that from trust in the truth of one proposition, we can infer trust in the truth of other propositions that are related by topic.

Second, we extend the logic with questions. In this way, we can express that informs are explicitly asked for, or else are implicitly considered relevant by an agent. There are two kinds of trust inference principles. We might say that by selecting another agent to ask a question, you indicate that you will trust this other agent. Thus, questions imply trust. On the other hand, questions may be asked strategically. In our running example the agent deliberately asked for a question with a known answer, in order to infer if the replying agent could be trusted on propositions of a related topic.

An important question concerns the applicability of trust principles. We have already seen two alternative principles regarding trust and questions. It also seems reasonable to restrict the trust derivation axiom to situations in which the agent is relatively ignorant. In an exam situation, the teacher knows the answers to all the questions he asks. But a correct answer to the first question will not necessarily make the teacher trust the student about the answers to the remaining questions. This just shows that the social context in which trust is applied, needs to be modeled very carefully.

There are several important properties of trust which remain undiscussed. The logic does not capture the element of risk. In the running example, trusting the web-service is risky, because the portfolio management of the agent depends on it. Note that without such a risk, the agent would not go through the trouble of testing the services with the question about exchange rates.

We briefly indicated how the logic might be further extended with requests and proposals. This however, would require a shift from an epistemic notion of trust, about beliefs, to a more practical notion of trust, about actions. We also discussed how the logic is related to more general transaction models of trust, which involve control mechanisms guaranteed by an institution. More research is needed to connect these models with work on institutional norms.

References

[1] Tan, Y.H., Thoen, W.: Formal aspects of a generic model of trust for electronic commerce. In: 33rd Hawaii International Conference on System Sciences (HICSS'00). (2000) p. 6006

[2] Liau, C.J.: Belief, information acquisition, and trust in multi-agent systems – a modal formulation. Artificial Intelligence **149** (2003) 31–60

[3] Gasquet, O., Herzig, A.: From classical to normal modal logics. In Wansing, H., ed.: Proof Theory of Modal Logics. Volume 2 of Applied Logic Series. Kluwer (1996) 293–311

[4] Kracht, M., Wolter, F.: Normal monomodal logics can simulate all others. Journal of Symbolic Logic **64** (1999) 99–138

[5] Goffman, E.: Strategic interaction. University of Pennsylvania Press, Pennsylvania (1969)

[6] Demolombe, R.: To trust information sources: a proposal for a modal logical framework. In Castelfranchi, C., Tan, Y.H., eds.: Trust and Deception in Virtual Societies. Kluwer (2001) 111 – 124

[7] Demolombe, R., Jones, A.: On sentences of the kind "sentence 'p' is about topic t". In: Logic, language and reasoning: Essays in Honour of Dov Gabbay. Kluwer Academic, Dordrecht (1999) 115–133
[8] Herzig, A., Longin, D.: Belief dynamics in cooperative dialogues. Journal of Semantics **17** (2000) 91–118
[9] Janin, D., Walukiewicz, I.: Automata for the modal mu-calculus and related results. In: Proceedings of the 20th International Symposium on Mathematical Foundations of Computer Science (MFCS'95). LNCS 969, Springer Verlag (1995) 552 – 562
[10] Groenendijk, J., Stokhof, M.: Questions. In Van Benthem, J., Ter Meulen, A., eds.: Handbook of Logic and Language. North-Holland, Elsevier (1996) 1055–1124
[11] van Kuppevelt, J.: Discourse structure, topicality and questioning. Journal of Linguistics **31** (1995) 109–149
[12] Horty, J.F.: Agency and Deontic Logic. Oxford University Press (2001)
[13] Tan, Y.H., Thoen, W.: An outline of a trust model for electronic commerce. Applied Artificial Intelligence **14** (2000) 849–862
[14] Mayer, R., Davis, J., Schoorman, F.: An integrative model of organizational trust. Academy of Management Review **20** (1995) 709–734
[15] Gambetta, D.: Can we trust trust? In: Trust. Basil Blackwell, New York (1988) 213–237
[16] Grosz, B.J., Sidner, C.L.: Attentions, intentions and the structure of discourse. Computational Linguistics **12** (1986) 175–204
[17] Huibers, T.: An Axiomatic Theory for Information Retrieval. PhD thesis, Utrecht University (1996)
[18] Burrows, M., Abadi, M., Needham, R.: A logic of authentication. ACM Transactions on Computer Systems **8** (1990) 18–36
[19] Foster, I., Kesselman, C., Tuecke, S.: The anatomy of the Grid: Enabling scalable virtual organizations. .International Journal of High Performance Computing Applications **15** (2001) 200–222
[20] Carbone, M., Nielsen, M., Sassone, V.: A formal model for trust in dynamic networks. In: International Conference on Software Engineering and Formal Methods (SEFM'03), IEEE (2003) 54–63
[21] Jones, A., Firozabadi, B.S.: On the characterisation of a trusting agent - aspects of a formal approach. In Castelfranci, C., Tan, Y., eds.: Trust and Deception in Virtual Societies. Kluwer Academic Publishers (2001) 157–168
[22] Blaze, M., Feigenbaum, J., Lacy, J.: Decentralized trust management. In: IEEE Symposium on Security and Privacy. IEEE (1996) 164–173
[23] Dellarocas, C.: The digitization of word-of-mouth: Promise and challenges of online feedback mechanisms. Management Scienc **49** (2004) 1407–1424